「정보시스템연구」 제32권 제3호
한국정보시스템학회
2023년 9월, pp. 01～17

http://dx.doi.org/10.5859/KAIS.2023.32.3.1

Multimodal Sentiment Analysis for Investigating User Satisfaction*

Hwang, Gyo Yeob** · Song, Zi Han*** · Park, Byung Kwon****

〈목 차〉

Ⅰ. Introduction

The development of the Internet has led to explosive growth in data size and computer performance, causing significant changes in many fields. In the area of user satisfaction surveys, traditional methods (Duleba and Moslem, 2021; Hayes, 2008) involve collecting and recording opinions through questionnaires, sample surveys, interviews, and focus groups. However, these survey methods are increasingly ineffective in dealing with the growing volume and diversity of data. For instance, users often upload large amounts of video data to social networking sites, which may contain valuable feedback on their opinions and needs for specific products. Obtaining this kind of information is crucial for accurately surveying user satisfaction, yet it is challenging to do so through traditional

* This work was supported by the Dong-A University research fund.
** Ph.D. Candidate, Department of Management Information System, Dong-A University, gyoyeob@gmail.com(First author)
*** Ph.D. Student, Department of Management Information System, Dong-A University, s2403353662@gmail.com
**** Professor, Department of Management Information System, Dong-A University, bpark@dau.ac.kr.(Corresponding author)

methods.

Numerous methods have been proposed for analyzing vast amounts of unstructured Internet data to explore user satisfaction. Current research mainly focuses on textual data, such as user reviews on shopping websites and articles on social media. One study (Hasson et al., 2019) investigated the use of social media data as an alternative to traditional customer satisfaction surveys. The authors compared customer feedback obtained through survey responses with feedback from Twitter and found that Twitter provided more information. In another study (Li et al., 2013), the authors used text-processing techniques to analyze 42,668 hotel reviews of 774 star-rated hotels to identify factors that contribute to customer satisfaction in hotel establishments. This approach demonstrated that surveying user satisfaction based on Internet content can yield a significantly larger number of samples than traditional questionnaire surveys, without significantly increasing costs.

While there are numerous comment-based customer satisfaction studies, only a few have explored user satisfaction through video and audio data. One study (Seng and Ang, 2018) proposed an audio-visual sentiment extraction system for video conferencing, which converts detected customer sentiment into a customer satisfaction score. In another study (Naas and Sigg, 2020), the authors reported an average accuracy of 78% in detecting user satisfaction using audio and video models. However, these studies did not consider the textual data contained in the videos, limiting system performance. In recent years, online video sites have experienced significant growth (Schwemmer and Ziewiecki, 2018). People are increasingly sharing their life experiences, including product experiences, through videos. These videos can influence potential customer purchase decisions. Therefore, it is crucial to extend user satisfaction surveys to online video sites.

Multimodal sentiment analysis has gained attention due to its high accuracy and broad applicability (Huddar et al., 2019; Morency et al., 2011; Soleymani et al., 2017). Unlike traditional text-based sentiment analysis, multimodal sentiment analysis integrates multiple modalities, including audio-visual data, to detect emotions. This approach utilizes deep learning techniques to detect voice tones and facial expressions, enabling the identification of emotional tendencies. For example, sentiment analysis can be performed on video content by combining facial expression emotion, tone emotion, and text emotion (Morency et al., 2011; Wollmer et al., 2013). Moreover, it utilizes emotion recognition and contextual inference to determine the underlying polarity and range of individual emotions. Therefore, multimodal sentiment analysis has considerable potential in online video sentiment analysis. However, the existing research on the use of multimodal

sentiment analysis in product customer satisfaction surveys is limited. Extracting emotional tendency from the vast array of user-uploaded online videos can contribute significantly to a company's understanding of product-related customer satisfaction. This user satisfaction derived from online videos offers invaluable insights for businesses, enabling them to enhance user experience, product or service quality, and overall customer loyalty.

This study aims to implement user satisfaction surveys in online videos using multimodal sentiment analysis techniques. Specifically, we utilize multimodal sentiment analysis to analyze the user satisfaction of iPhone series products and Samsung series products through a case study approach. We searched for a total of 528 YouTube videos containing the keywords iPhone SE, iPhone 11, iPhone 11 Pro, iPhone 12, iPhone 12 Pro, iPhone 13, iPhone 13 Pro, Samsung S21, Samsung S21+, Samsung S22, and Samsung S22+. We analyzed the emotional tendencies of the videos from different sources. To validate the performance of the model, we manually annotated the emotional inclinations of the data. Finally, we discussed the user product satisfaction for each product in the case study. The contributions of this research can be summarized as follows:

(1) To the best of our knowledge, this is the first study to employ multimodal sentiment analysis techniques to analyze product user satisfaction in online videos.

(2) The research explores the importance of multimodal data sources for the accuracy of sentiment analysis in online videos.

(3) The performance of the model is analyzed and discussed across a total of 528 online videos for various iPhone and Samsung products.

The paper is organized as follows: In Section 2, we review the current research on multimodal sentiment analysis. In Section 3, we describe the specific machine learning techniques used in our study. In Section 4, we analyze the sentiment data results from different sources and explore the correlations among them. In Section 5, we present our discussion and conclusions.

Ⅱ. Literature Review

2.1 Multimodal sentiment analysis

Multimodal sentiment analysis is a technological advancement that aims to accurately comprehend emotional tendencies based on information from different modalities, such as text, voice, and images. Traditional sentiment analysis that relies solely on textual information is not as accurate as multimodal sentiment analysis (Yadollahi et al., 2018). Due to its potential to enhance understanding of user sentiment and its role in improving the

overall user experience, multimodal sentiment analysis has become an active area of research in the field of online video platforms (Huddar et al., 2019; Soleymani et al., 2017).

Several studies have examined the effectiveness of multimodal sentiment analysis in the context of online video platforms. For instance, Xu et al. (Xu et al., 2020) employed deep learning techniques to analyze video content and discovered that a multimodal approach improved the accuracy of sentiment analysis by approximately 15% compared to using a single-input model. Similarly, another study (Li et al., 2020) utilized a multimodal approach to analyze the sentiment of YouTube posts about emotional car reviews and discovered that the combination of various models significantly enhanced the accuracy of sentiment analysis.

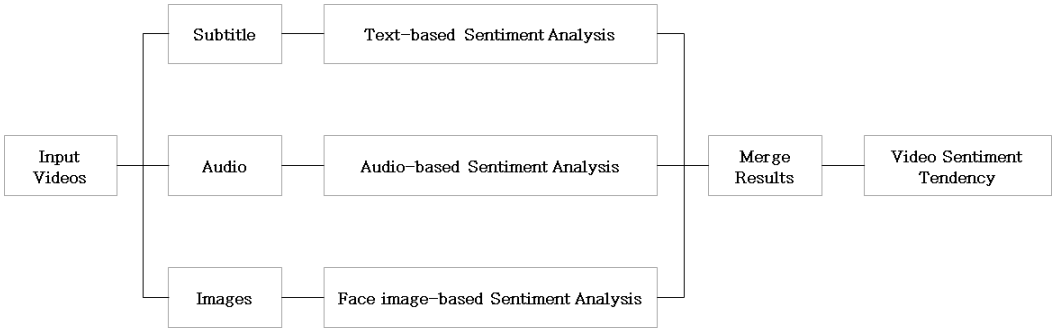
In another study (Perez Rosas et al., 2013), a combination of text-based and visual-based models was employed to analyze the sentiment of video comments on Spanish videos. The authors discovered that characteristics such as smiles and pauses are typical for positive videos, whereas higher voice intensity is more commonly associated with negative videos.

Taken together, these studies indicate that multimodal sentiment analysis is a powerful technique for understanding user sentiment towards videos and channels on online video platforms.

2.2 Online product user satisfaction survey

Online product satisfaction surveys have become an important tool for companies to understand and measure the satisfaction of their customers with their products (Mansor et al., 2018). Companies can easily reach a large number of customers through online surveys and can receive the results in real time. To get better user feedback on products, companies are trying to use a multi-channel approach to reach customers (Madaleno et al., 2007), using a combination of email, text message, and social media to invite customers to take surveys. Currently, artificial intelligence technology is used to analyze survey results and provide more personalized survey questions (Prentice et al., 2020).

However, at present, the main method of obtaining product feedback from users is still questionnaires or telephone interviews. Compared with these methods that require users to actively provide product feedback, extracting product feedback information from user-uploaded comments, videos and other data has higher flexibility and lower cost. Previous research has illustrated the importance of investigating user comments (Liu et al., 2019). Compared to product reviews, user-generated videos may be seen by more people and have higher influence. Investigating users' perceptions of products in these videos can be critical to



<Figure 1> Process pipeline of the multimodal sentiment analysis

corporate decision-making. However, due to the unstructured features of video data, few studies have attempted to extract user feedback from it. Therefore, this study aims to implement user satisfaction surveys in online videos using multimodal sentiment analysis techniques.

Ⅲ. Multimodal Sentiment Anlaysis for YouTube Videos

We conducted a search for a total of 336 YouTube videos using keywords such as iPhone SE, iPhone 11, iPhone 11 Pro, iPhone 12, iPhone 12 Pro, iPhone 13, and iPhone 13 Pro. Each video in our dataset includes both audio and subtitles. Our analysis process involved segmenting the videos into image sequences to enable facial emotion classification. Additionally, we extracted the subtitles and audio for text sentiment classification and audio sentiment classification, respectively. Figure 1 illustrates the pipeline of our multimodal sentiment analysis process.

3.1 Text Sentiment Analysis

Since 2000, sentiment analysis has emerged as a prominent area of research within natural language processing. Text sentiment analysis methods can be broadly categorized into two types: the lexicon-based approach and the machine learning-based approach. The lexicon-based approach employs a sentiment dictionary to calculate the sentiment score of a given text based on the occurrence of sentiment words within the text, and to thereby determine the overall sentiment tendency. These sentiment dictionaries can be manually or semi-manually constructed. On the other hand, the machine learning-based approach automatically classifies reviews by leveraging training data. Machine learning techniques are utilized to develop an algorithm and to build a model by employing feature selection and learning from labeled training datasets (Hu and Liu, 2006). Well-

known machine learning methods for sentiment classification include the Naïve Bayes Classifier, Support Vector Machine (SVM), and Random Forest.

The TextBlob (“TextBlob Homepage”) library is a widely used open-source library for performing sentiment analysis on textual data. It provides a simple interface for performing sentiment analysis and other natural language processing tasks. TextBlob uses a lexicon-based approach to sentiment analysis, where each word in a sentence is assigned a sentiment score based on its polarity. The polarity score for the entire sentence is then calculated by aggregating the scores of all the words in the sentence. Using TextBlob allowed us to efficiently analyze the sentiment of a large amount of textual data without the need for manual annotation or feature engineering. Additionally, TextBlob has been shown to produce reliable results in previous studies and has been widely used in sentiment analysis research (Ahuja and Dubey, 2017).

In our study, we utilized the TextBlob library to classify the sentiment of each sentence in the video subtitles. Specifically, we classified each sentence as negative, neutral, or positive based on its overall sentiment score. The sentiment scores were obtained by using the pre-trained sentiment analyzer provided by the TextBlob library. The analyzer utilizes a pre-built sentiment lexicon that contains polarity scores for a large number of words. The polarity scores range from -1 (extremely negative) to 1 (extremely positive), with 0 indicating a neutral sentiment.

3.2 Image Sentiment Analysis

Since the sentiment of images is more abstract and subjective, the task of image sentiment analysis is more complex than text sentiment analysis. Image sentiment analysis involves the analysis of visual features such as color, texture, and shape to understand the emotions and feelings that the image conveys. However, the interpretation of these visual features can vary from person to person, making image sentiment analysis a more challenging task than text sentiment analysis.

The DeepFace library (Serengil and Ozpinar, 2021) is a popular and powerful tool for facial emotion recognition, making it well-suited for addressing the challenge of image sentiment analysis. The library utilizes deep learning techniques, which allow it to analyze images at a very detailed level, identifying subtle nuances in facial expressions that may not be obvious to the human eye. Furthermore, DeepFace is trained on a large dataset of faces, allowing it to recognize emotions in a variety of contexts and with a high degree of accuracy. This is important for image sentiment analysis, as emotions can be influenced by a wide range of factors, such as lighting, facial angle, and individual differences

in expressions.

This research focuses on the emotional tendencies of the characters in the short video image, therefore only the emotional information of the characters is extracted in the image sentiment analysis. We divided the videos into frames and use DeepFace to analyze the sentiment tendency. The DeepFace divides the characters' expressions into the following 7 types: angry, fear, neutral, sad, disgust, happy, surprise.

Because the emotional tendency obtained from facial expression recognition can only represent the emotional tendency of a person in a single image (one frame in video). Therefore, to get the emotional tendency of the overall video, we merge all the emotional tendency that appear in the entire video. To do this we assign anger, fear, sad, and disgust as negative emotions (-1), happy and surprise as positive emotions (1).

3.3 Audio Sentiment Model

Audio sentiment analysis is a technique used to analyze the sentiment or emotional tone expressed in audio data, such as speech and music. The primary objective of audio sentiment analysis is to classify the emotional state of audio data as positive, negative, or neutral. However, due to the limited availability of publicly accessible audio emotion classification models, developing an accurate audio emotion classification model is a challenging task.

In this study, we have trained an audio emotion classification model based on a publicly available dataset (“Audio Speech Sentiment”) consisting of 360 annotated audio clips that are classified into three categories: positive, negative, and neutral. We employed a deep neural network architecture containing 12 layers and a total of 557,486 parameters to train our model. The feature data of audio clips were extracted using functions provided by the librosa package (“librosa”). The extracted features included zero-crossing rate, chromogram, Mel-frequency cepstral coefficients, root-mean-square (RMS) value for each frame, and Mel-scaled spectrogram.

After completing the training phase, we evaluated the performance of our model on the test set. The classification accuracy of our model was found to be 98%. This indicates that our model is highly accurate in identifying the emotional state of audio data.

IV. Results

4.1 Data Acquisition

We collected publicly available videos from YouTube using web scraping techniques. The search keywords included iPhone SE, iPhone 11, iPhone 11 Pro, iPhone 12, iPhone 12 Pro,

iPhone 13, iPhone 13 Pro, Samsung S21, Samsung S21+, Samsung S22, and Samsung S22+. For each keyword, we gathered 100 videos from the search results. Subsequently, we manually filtered the videos to remove duplicates and those unrelated to product reviews. The final dataset comprised 48 videos for each keyword, resulting in a total of 528 valid videos for analysis. To construct the dataset, we manually annotated video content by classifying it according to sentiment categories. Specifically, the criterion for manual annotation is the sentiment orientation of the current video for the specified keywords (positive, neutral, and negative).

The classification results of each model were evaluated by comparing them to these manually constructed labels.

4.2 Evaluation Metrics

We evaluate the performance of our model using four key metrics: accuracy, F1-score, precision, and recall, as shown in Table 1. Since we have divided the videos into three categories (positive, neutral, and negative), we employ the Macro method to calculate the average precision, recall, and F1-score.

Accuracy: Accuracy is the ratio of the correctly predicted instances to the total instances in the dataset. It is calculated as follows:

(1) Accuracy

= (TP + TN) / (TP + FP + FN + TN)

Macro F1-score: The macro F1-score takes into account both false positives and false negatives, making it a more robust measure of model performance, especially when dealing with imbalanced datasets. It calculates the F1-score for each class individually and then takes the average, giving equal weight to each class. It is calculated as follows:

(2) Macro F1-Score

= 2 x (Macro Precison x Macro recall) / (Macro Precison + Macro Recall)

Macro Precision: Macro precision is the average of the precision scores for each class. Precision is the ratio of true positives to the sum of true positives and false positives. It is calculated as follows:

(3) Macro Precision = TP / (TP + FP)

Macro Recall: Macro recall is the average of the recall scores for each class. Recall, also known as sensitivity or true positive rate, is the ratio of true positives to the sum of true positives and false negatives. It is calculated as

Table 1 : The Confusion Matrix

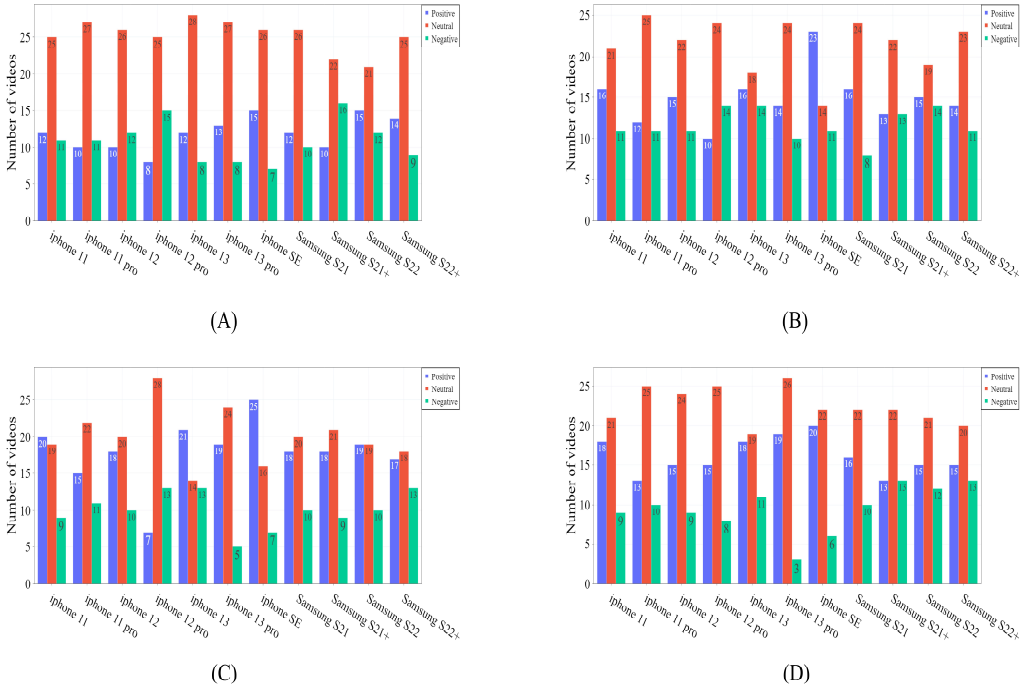
		Actual	
		Positive	**Negative**
Predicted	**Positive**	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

follows:

(4) Macro Recall = TP / (TP + FN)

4.3 Experimental Results

Figure 2.A, Figure 2.B, and Figure 2.C show the results of multimodal sentiment analysis based on the results of multimodal sentiment analysis based the results of multimodal the results of multimodal sentiment analysis based on video, audio, and subtitle information. Figure 2.D depicts the outcomes of sentiment analysis that are derived from a combination of subtitles, videos, and tone of voice. The sentiment analysis results obtained from these three sources are merged using an averaging method. The Y axis represents the number of videos corresponding to the emotional tendency.



(A) (B)

(C) (D)

<Figure 2> (A) Results of sentiment analysis based on subtitles. (B) Results of sentiment analysis based on human face occurred in videos. (C) Results of sentiment analysis based on tone of voice. (D) Results of sentiment analysis based on the combination result of subtitles, videos, and tone of voice.

<Table 2> The performance of sentiment analysis for seven different iPhone models (SE, 11, 11 Pro, 12, 12 Pro, 13, and 13 Pro) and four Samsung models (S21, S21+, S22, S22+) based on four different models' result: image, tone, subtitle, and combination result. The positions of the highest accuracy, F1-score, precision, and recall of the current row are marked in bold font.

Product	Image				Tone			
	Accuracy	F1-Score	Precision	Recall	Accuracy	F1-Score	Precision	Recall
iPhone SE	0.6	0.59	0.58	0.6	0.55	0.51	0.44	0.6
iPhone 11	0.57	0.55	0.55	0.56	0.49	0.46	0.45	0.47
iPhone 11 pro	0.62	**0.60**	**0.59**	**0.62**	0.47	0.49	0.46	0.52
iPhone 12	0.55	0.54	0.54	0.54	0.52	0.42	0.41	0.43
iPhone 12 pro	0.61	0.59	**0.59**	0.6	0.56	0.51	0.48	0.54
iPhone 13	0.6	0.60	0.6	0.61	0.54	0.54	0.52	0.56
iPhone 13 pro	0.54	0.54	0.54	0.55	0.5	0.48	0.48	0.49
Samsung S21	0.62	0.56	0.56	0.57	0.55	0.54	0.54	0.55
Samsung S21+	0.66	**0.66**	**0.66**	**0.67**	0.51	0.51	0.49	0.54
Samsung S22	0.64	0.54	0.54	0.55	0.53	0.53	0.52	0.55
Samsung S22+	0.63	0.57	0.56	0.58	0.54	0.46	0.43	0.5
Average	0.60	0.58	0.57	0.59	0.52	0.50	0.47	0.52
Product	Subtitle				Combination			
	Accuracy	F1-Score	Precision	Recall	Accuracy	F1-Score	Precision	Recall
iPhone SE	0.65	**0.66**	**0.65**	**0.68**	**0.67**	0.62	0.6	0.65
iPhone 11	0.63	0.49	0.48	0.51	**0.65**	**0.65**	**0.65**	**0.66**
iPhone 11 pro	0.6	0.56	0.57	0.56	**0.66**	0.54	0.53	0.56
iPhone 12	0.63	**0.59**	**0.58**	**0.6**	**0.63**	0.56	0.52	0.6
iPhone 12 pro	0.61	0.59	0.58	0.61	**0.65**	**0.60**	**0.59**	**0.62**
iPhone 13	0.67	0.65	0.64	0.66	**0.72**	**0.74**	**0.69**	**0.79**
iPhone 13 pro	**0.66**	0.61	0.6	0.62	0.65	**0.66**	**0.63**	**0.69**
Samsung S21	0.66	**0.64**	**0.63**	**0.65**	**0.67**	0.62	0.61	0.64
Samsung S21+	0.65	0.60	0.57	0.63	**0.7**	0.60	0.58	0.63
Samsung S22	0.63	0.55	0.54	0.56	**0.69**	**0.66**	**0.65**	**0.67**
Samsung S22+	0.62	0.55	0.54	0.57	**0.68**	**0.64**	**0.64**	**0.65**
Average	0.64	0.59	0.58	0.60	**0.67**	**0.63**	**0.61**	**0.65**

The results of sentiment analysis based on different sources have significant differences. In the classification results based on subtitle information, the proportion of neutral is the highest, in the results based on video facial expression recognition, the proportion of negative is relatively the highest, and in the results of classification based on tone, the proportion of positive videos is relatively the highest. More specifically, among the three results, the number of videos classified as positive have similar trends. The number of positive videos in the pro series or plus series is different from that of the normal series, and the number of positive videos in the iPhone SE is relatively the highest. Because the pro series or plus series of smartphone are relatively more powerful than the general products of the same series. Therefore, it can be considered that the reason for this difference is the price. This reveals the impact of price on user product satisfaction, the moSre affordable the price, the more user product satisfaction. Correspondingly, there are more videos expressing positive emotions.

The results of the analysis demonstrated that the accuracy of sentiment analysis varied depending on the data source and the smartphone model. Overall, the combination model yielded the most optimal performance, followed by the subtitle-based model. In terms of accuracy, the combination model exhibited superior performance across all products. Furthermore, the combination model demonstrated the best comprehensive classification performance (F1-Score) for half of the smartphone products. It is noteworthy that the tone-based and image-based models yielded poor performance in all cases, indicating that relying solely on tone and facial expression information is insufficient for extracting accurate emotional orientation. This evidence supports the notion that sentiment classification, derived from the integration of multimodal data analysis, is more reliable.

V. Discussion

The present study investigated the effectiveness of different data sources for sentiment analysis of iPhone and Samsung product video reviews, including image, tone, and subtitle information. The results obtained from the analysis showed that the proportion of videos varied depending on the data source, with the highest proportion of neutral videos among the results. Moreover, the number of positive videos was found to be related to the price of the smartphone model, with the more affordable models having more videos expressing positive emotions. This suggests that price is an important factor affecting user satisfaction, and that users tend to express more positive emotions when they are satisfied with the price of a product.

In addition, the accuracy of sentiment analysis varied depending on the data source and smartphone model. We found that the classification performance of the combination model outperformed other models using a single data source. This indicates that the multimodal approach to sentiment analysis is more reliable than using a single data source.

Because using information from a single data source may limit the scope of sentiment analysis, as it does not consider the full range of cues that convey emotional information. In contrast, the multimodal approach presented in this paper incorporates visual and auditory cues, which can provide a more comprehensive understanding of user sentiment. This is particularly important in the context of online videos, where non-textual elements such as images and tone often play a crucial role in conveying emotions and opinions.

Furthermore, among the three models based on a single data source, the performance metrics of the subtitle-based model and the image-based model are significantly higher than those of the tone-based model. This reveals that images and subtitles may reflect more emotional information. This is consistent with previous studies on multimodal sentiment analysis (Chen and Li, 2020). The accuracy of tone-based model was relatively low, which is likely due to the fact that tone of voice is easily affected by various factors such as background noise, speaker characteristics, and speech content (Cui et al., 2019).

Compared with the traditional questionnaire-based user satisfaction survey method (Khan et al., 2015), the proposed multimodal analysis method for online video has the following three advantages:

(1) Real-time feedback: Online videos can provide real-time, unfiltered insights into user satisfaction, whereas questionnaires may suffer from response biases or inaccuracies due to self-reporting.

(2) Scalability and richer data: Analyzing user sentiment through online videos using automated techniques can be more scalable than administering and processing questionnaires, as it can handle large volumes of data more efficiently.

(3) Cultural and linguistic diversity: Multimodal sentiment analysis can potentially capture user satisfaction across different languages and cultural contexts, as it relies on multiple cues rather than being solely dependent on text-based information.

However, if the analysis target is more detailed user satisfaction, considering aspects such as product price, convenience, and product information, rather than the overall emotional tendency, this method still has some limitations. Therefore, future research should try to conduct a more in-depth analysis of the video content, such as judging the attitude of the speaker in the video to each subdivided topic.

The results of this study show that using a multimodal sentiment analysis method can clearly reveal the emotional orientation of online videos to target products. These results have important implications for product design and marketing strategies. For instance, companies can use sentiment analysis to identify the emotional responses of consumers to their products and adjust their marketing strategies accordingly. By considering the impact of different factors on user satisfaction, such as price, companies can design products that meet the needs and preferences of their target audience. Overall, the study provides valuable insights into the effectiveness of different data sources for sentiment analysis of product reviews and highlights the importance of considering multiple factors when evaluating user satisfaction.

VI. Conclusion

This study investigated the sentiment analysis of iPhone and Samsung smartphone review videos by using multimodal data analysis, including video, audio, and subtitle information. The results showed that the sentiment analysis accuracy varied depending on the data source and smartphone model. The research reveals that the combination model integrating multiple data sources showed the most superior performance, indicating that the sentiment classification obtained by combining multimodal data analysis is more reliable. The findings also revealed that the price of smartphone products has a significant impact on user satisfaction, with more affordable products resulting in higher levels of user satisfaction and more positive emotional expression in review videos. This study provides valuable insights into the potential applications of multimodal data analysis in sentiment analysis and highlights the importance of considering multiple sources of information in analyzing sentiment. Future research can build upon these findings by exploring more advanced multimodal analysis techniques and investigating the sentiment analysis of other types of products and services.

References

Ahuja, S. and Dubey, G., Clustering and sentiment analysis on Twitter data. 2017 2nd International Conference on Telecommunication and Networks (TEL-NET), IEEE, 2017, pp. 1-5. Audio Speech Sentiment.

Chen, M. and Li, X. SWAFN: Sentimental Words Aware Fusion Network for Multimodal Sentiment Analysis. Proceedings of the 28th International Conference on Computational

Linguistics, International Committee on Computational Linguistics, Stroudsburg, PA, USA, 2020, pp. 1067-1077.

Cui, Z., Qiu, Q., Yin, C., Yu, J., Wu, Z. and Deng, A., "A Barrage Sentiment Analysis Scheme Based on Expression and Tone", IEEE Access, Vol. 7, 2109, pp. 180324-180335.

Duleba, S. and Moslem, S., "User Satisfaction Survey on Public Transport by a New PAHP Based Model", Applied Sciences, Vol. 11, No. 21, 2021, p. 10256.

Hasson, S.G., Piorkowski, J. and McCulloh, I., Social media as a main source of customer feedback. Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM, New York, NY, USA, 2109, pp. 829-832.

Hayes, B.E., Measuring customer satisfaction and loyalty: survey design, use, and statistical analysis methods, Quality Press, 2008.

Hu, M. and Liu, B., "Opinion extraction and summarization on the Web", Proceedings of the National Conference on Artificial Intelligence, Vol. 2, 2006, pp. 1621-1624.

Huddar, M.G., Sannakki, S.S. and Rajpurohit, V.S., "A Survey of Computational Approaches and Challenges in Multimodal Sentiment Analysis", International Journal of Computer Sciences and Engineering, Vol. 7, No. 1, 2019, pp. 876-883.

Khan, S.A., Liang, Y. and Shahzad, S., "An Empirical Study of Perceived Factors Affecting Customer Satisfaction to Re-Purchase Intention in Online Stores in China:" Journal of Service Science and Management, Vol. 08, No. 03, 2015, pp. 291-305.

Li, H., Ye, Q. and Law, R., "Determinants of Customer Satisfaction in the Hotel Industry: An Application of Online Review Analysis", Asia Pacific Journal of Tourism Research, Vol. 18, No. 7, 2013, pp. 784-802.

Li, R., Zhao, J., Hu, J., Guo, S. and Jin, Q., Multi-modal Fusion for Video Sentiment Analysis. Proceedings of the 1st International on Multimodal Sentiment Analysis in Real-life Media Challenge and Workshop, ACM, New York, NY, USA, 2020, pp. 19-25.

librosa. Available at: https://librosa.org/doc/latest/index.html (Accessed: February 7, 2023).

Lin, C.-C., Wu, H.-Y. and Chang, Y.-F., "The critical factors impact on online customer satisfaction", Procedia Computer Science,Elsevier Vol. 3, 2011, pp. 276-281.

Liu, Y., Jiang, C. and Zhao, H., "Assessing product competitive advantages from

the perspective of customers by mining user-generated content on social media:" Decision Support Systems, Vol. 123, 2019, p. 113079.

Madaleno, R., Wilson, H. and Palmer, R., "Determinants of Customer Satisfaction in a Multi-Channel B2B Environment", Total Quality Management & Business Excellence, Vol. 18, No. 8, 2007, pp. 915-925.

Mansor, S.N., Mostafa, S.A., Mustapha, A. and Darman, R., An Emotional Agent for the Analysis of Customer Satisfaction Surveys., 2018 International Symposium on Agent, Multi-Agent Systems and Robotics (ISAMSR), IEEE, 2018, pp. 1 -6.

Morency, L.-P., Mihalcea, R. and Doshi, P., Towards multimodal sentiment analysis. Proceedings of the 13th international conference on multimodal interfaces - ICMI '11, ACM Press, New York, New York, USA, 2011, p. 169.

Naas, S.A. and Sigg, S., "Real-time emotion recognition for sales:" Proceedings - 2020 16th International Conference on Mobility, Sensing and Networking, MSN 2020, pp. 584-591.

Perez Rosas, V., Mihalcea, R. and Morency, L.-P., "Multimodal Sentiment Analysis of Spanish Online Videos", IEEE Intelligent Systems, Vol. 28, No. 3, 2013, pp. 38-45.

Prentice, C., Dominique Lopes, S. and Wang, X., "The impact of artificial intelligence and employee service quality on customer satisfaction and loyalty", Journal of Hospitality Marketing & Management, Vol. 29, No. 7, 2020, pp. 739-756.

Schwemmer, C. and Ziewiecki, S., "Social Media Sellout: The Increasing Role of Product Promotion on YouTube", Social Media + Society, Vol. 4, No. 3, 2018, p. 205630511878672.

Seng, K.P. and Ang, L.M., "Video analytics for customer emotion and satisfaction at contact centers", IEEE Transactions on Human-Machine Systems, Vol. 48, No. 3, 2018, pp. 266-278.

Serengil, S.I. and Ozpinar, A., HyperExtended LightFace: A Facial Attribute Analysis Framework, 2021 International Conference on Engineering and Emerging Technologies (ICEET), IEEE, 2021, pp. 1-4.

Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S.F. and Pantic, M., "A survey of multimodal sentiment analysis", Image and Vision Computing,Elsevier B.V. Vol. 65, 2107, pp. 3-14.

TextBlob Homepage. Available at: https:// textblob.readthedocs.io/en/dev/(Accessed : June 9, 2022).

Wollmer, M., Weninger, F., Knaup, T., Schuller, B., Sun, C., Sagae, K. and

Morency, L.-P., “YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context”, IEEE Intelligent Systems, Vol. 28, No. 3, 2013, pp. 46-53.

Xu, G., Li, W. and Liu, J., “A social emotion classification approach using multi-model fusion”, Future Generation Computer Systems, Vol. 102, 2020, pp. 347-356.

Yadollahi, A., Shahraki, A.G. and Zaiane, O.R., “Current State of Text Sentiment Analysis from Opinion to Emotion Mining”, ACM Computing Surveys, Vol. 50, No. 2, 2018, pp. 1-33.

황 교 엽 (Hwang, Gyo Yeob)

동국대학교 경영학사, 동대학원의 경영전문대학원에서 석사학위를 취득하였다. 현재 동아대학교 경영정보학과 박사과정 수료하였으며, 주요 관심분야는 정보기술, 기업경영, 머신러닝 및 제조업의 사무·산업자동화이다.

송 쯔 한 (Song, Zi Han)

중국장시과학기술대학 공학학사와 동아대학교 경영학 석사학위를 취득하였다. 현재 동아대학교 경영정보학과에 재학중이다. 주요 관심 분야는 머신러닝, 데이터분석이다.

박 병 권 (Park, Byung Kwon)

서울대학교 산업공학과에서 공학사와 KAIST 경영과학 석사, KAIST 전산학과에서 공학 박사학위를 취득하였다. 현재 동아대학교 경영정보학과 교수로 재직하고 있으며, 주요 관심 분야는 비즈니스 인텔리전스 시스템, 제조업의 스마트 트랜스폼, 스마트 컨테이너와 스마트 해운물류 등이다.

<Abstract>

Multimodal Sentiment Analysis for Investing User Satisfaction

Hwang, Gyo Yeob · Song, Zi Han · Park, Byung Kwon

Purpose

The proliferation of data on the internet has created a need for innovative methods to analyze user satisfaction data. Traditional survey methods are becoming inadequate in dealing with the increasing volume and diversity of data, and new methods using unstructured internet data are being explored. While numerous comment-based user satisfaction studies have been conducted, only a few have explored user satisfaction through video and audio data. Multimodal sentiment analysis, which integrates multiple modalities, has gained attention due to its high accuracy and broad applicability.

Design/methodology/approach

This study uses multimodal sentiment analysis to analyze user satisfaction of iPhone and Samsung products through online videos. The research reveals that the combination model integrating multiple data sources showed the most superior performance.

Findings

The findings also indicate that price is a crucial factor influencing user satisfaction, and users tend to exhibit more positive emotions when content with a product's price. The study highlights the importance of considering multiple factors when evaluating user satisfaction and provides valuable insights into the effectiveness of different data sources for sentiment analysis of product reviews.

Keyword: Multimodal sentiment analysis, Machine Learning, User satisfaction surveys, Online video sentiment analysis.

* 이 논문은 2023년 3월 27일 접수, 2023년 4월 11일 1차 심사, 2023년 7월 12일 2차 심사, 2023년 7월 18일 게재 확정되었습니다.